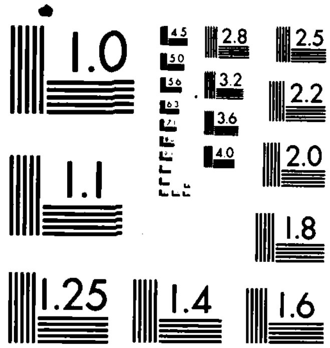
AD-A162 333 ASYMPTOTIC GLOBAL BEHAVIOR FOR STOCHASTIC APPROXIMATIONS AND DIFFUSIONS W (U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE 1/1

UNCLASSIFIED H J KUSHNER APR 85 AFOSR-TR-85-1022 F/G 12/1 NL

END

FILMED

DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(2)

AFOSR-TR- 85-1022

		Approved for public rel[...] unlimited	
[...]WNGRADING SCHEDULE			
[...]ATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION RE[...] AFOSR-TR- 8[...]	
[...] ORGANIZATION [...] for Dynamical	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANI[...] AFOSR	
[...] ZIP Code) [...]ied Mathematics 02912		7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 2033[...]	

DTIC ELECTE DEC 0 9 198[...] S D

AD-A162 333

Lefschetz Center for Dynamical Systems

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Division of Applied Mathematics   Brown University Providence RI 02912

Unclassifi̇ed

SECURITY CLASSIFICATION OF THIS PAGE

AD-A162 333

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY ---	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 85-1022
6a. NAME OF PERFORMING ORGANIZATION Lefschetz Center for Dynamical Systems 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR
6c. ADDRESS (City, State and ZIP Code) Division of Applied Mathematics Brown University Providence, RI 02912	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-81-0116
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A1	

11. TITLE (Include Security Classification) Asym· Global Behavior for Stochastic Approximations and Diffusions with Slowly Decreasing Noise Effects

12. PERSONAL AUTHOR(S)
H. J. Kushner

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Interim	FROM ______ TO ______	April 1985	35

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
XXXXXXXXXXXXX		

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Monte Carlo, stochastic approximation, large deviations simulated annealing, global function optimization from noisy samples

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
The asymptotic behavior of the systems $X_{n+1} = X_n + a_n b(X_n, \xi_n) + a_n \sigma(X_n)\nu_n$ and $dy = \bar{b}(y)dt + \sqrt{a(t)}\ \sigma(y)dw$ is studied, where $\{\nu_n\}$ is i.i.d. Gaussian, $\{\xi_n\}$ is a (correlated) bounded sequence of random variables and $a_n \approx A_0/\log(A_1 = n)$. Without $\{\xi_n\}$, such algorithms are versions of the 'simulated annealing' method for global optimization. When the objective function values can only be sampled via Monte Carlo, the discrete algorithm is a combination of stochastic approximation and simulated annealing. Our forms are appropriate. The $\{\nu_n\}$ are the 'annealing' variables, and $\{\xi_n\}$ is the sampling noise. For large A_0, a full asymptotic analysis is presented, via the theory of large deviations: Mean escape time (after arbitrary time n) from neighborhoods of stable sets of the (cont)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Brian W. Woodruff, Maj, USAF	(202)767-5027	NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

Block #19 cont.

algorithm, mean transition times (after arbitrary time n) from a neighborhood of one stable set to another, approximate asymptotic invariant measures, and location of the values of $\{X_n\}$ or $y(\cdot)$ the case where $Eb(x,\xi) = \bar{b}(x)$, and application to global function minimization via Monte Carlo methods.

(2)

ASYMPTOTIC GLOBAL BEHAVIOR FOR STOCHASTIC APPROXIMATIONS AND DIFFUSIONS WITH SLOWLY DECREASING NOISE EFFECTS: GLOBAL MINIMIZATION VIA MONTE CARLO

by

H. J. Kushner

April 1985 LCDS #85-7

ASYMPTOTIC GLOBAL BEHAVIOR FOR STOCHASTIC APPROXIMATIONS AND DIFFUSIONS WITH SLOWLY DECREASING NOISE EFFECTS: GLOBAL MINIMIZATION VIA MONTE CARLO

by

H. J. Kushner

Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

This research was supported by the Air Force Office of Scientific Research under Grant #AF AFOSR 81-0116-C; by the Army Research Office under grant #DAAG-29-84-K-0082; and by the Office of Naval Research under grant # N00014-84-C-0577.

ASYMPTOTIC GLOBAL BEHAVIOR FOR STOCHASTIC APPROXIMATION AND DIFFUSIONS WITH SLOWLY DECREASING NOISE EFFECTS: GLOBAL MINIMIZATION VIA MONTE CARLO

by

H. J. Kushner

ABSTRACT

The asymptotic behavior of the systems $X_{n+1} = X_n + a_n b(X_n, \xi_n) + a_n \sigma(X_n)\psi_n$ and $dy = \bar{b}(y)dt + \sqrt{a(t)}\,\sigma(y)dw$ is studied, where $\{\psi_n\}$ is i.i.d. Gaussian, $\{\xi_n\}$ is a (correlated) bounded sequence of random variables and $a_n \approx A_0/\log(A_1+n)$. Without $\{\xi_n\}$, such algorithms are versions of the 'simulated annealing' method for global optimization. When the objective function values can only be sampled via Monte Carlo, the discrete algorithm is a combination of stochastic approximation and simulated annealing. Our forms are appropriate. The $\{\psi_n\}$ are the 'annealing' variables, and $\{\xi_n\}$ is the sampling noise. For large A_0, a full asymptotic analysis is presented, via the theory of large deviations: Mean escape time (after arbitrary time n) from neighborhoods of stable sets of the algorithm, mean transition times (after arbitrary time n) from a neighborhood of one stable set to another, approximate asymptotic invariant measures, and location of the values of $\{X_n\}$ or $y(\cdot)$ the case where $Eb(x,\xi) = \bar{b}(x)$ is the (negative) of a gradient of a function $\bar{B}(x)$, and application to global function minimization via Monte Carlo methods.

Key words. Monte Carlo, stochastic approximation, large deviations, simulated annealing, global function optimization from noisy samples

1. Introduction

This document studies

~~We study~~ the asymptotic behavior of ~~the~~ a system involving a sequence of random variables. Keywords: Monte Carlo method, Stochastic approximation; algorithms; annealing.

$$X_{n+1} = X_n + a_n b(X_n,\xi_n) + a_n\sigma(X_n)\psi_n, \quad X \in R^r, \tag{1.1}$$

where $\{\xi_n\}$ is a sequence of bounded random variables, $\{\psi_n\}$ is a sequence of zero mean independent and identically distributed (i.i.d.) Gaussian random variables, the two sequences are mutually independent and $a_n = A_0/\log(n+A_1)$, $A_0 > 0$, $A_1 > 1$. The $\sigma(\cdot)$ and $b(\cdot,\xi)$ are Lipschitz continuous, uniformly in ξ, and $\sigma(\cdot)$ is bounded.

Such stochastic approximation algorithms are a Monte-Carlo version of the currently popular 'annealing' method for locating the global minimum of a function with many minima [1]-[3]: For example, let $\overline{B}(.)$ denote a continuously differentiable function and set $Eb(x,\xi_n) = \overline{b}(x) = -\overline{B}_x(x)$. Suppose that noise corrupted samples of $\overline{b}(x)$, namely $b(x,\xi)$, are available from an experiment or a simulation on a system whose 'mean' performance is $\overline{B}(x)$ at parameter value x. Then the algorithm $Y_{n+1} = Y_n + \alpha_n b(Y_n,\xi_n)$ is a standard form of a stochastic approximation method for locating a local zero of $\overline{b}(\cdot)$ or local minimum of $\overline{B}(\cdot)$ under appropriate conditions on $\{\alpha_n\}$. The $\sigma(x)\psi$ term might be added <u>artificially</u>, following the usual logic of the 'annealing' scheme, in order to force the sequence to jump around until it eventually 'settles' near a global minimum of $\overline{B}(\cdot)$. When only random samples $b(x,\xi)$ are available, the situation is much more complex than in the non-random sampling case. It is important to allow the $\{\xi_n\}$ to be correlated, since (a) many efficient Monte Carlo methods (e.g., antithetic variables) require correlated noise, or (b) simulations are often run on a continuously operating system, where the noise is inherently correlated.

The theory of large deviations [4], [5], [7], provides the appropriate methods, and allows us to obtain a fairly complete characterization of the asymptotic location of and behavior of the $\{X_n\}$.

If the rate of decrease of a_n is 'faster' than $O(1/\log n)$, then (under broad condition) $\{X_n\}$ will converge w.p.1, and not continue to 'try' to escape from the stable set of the algorithm in which it is trapped. In general, if $\sigma(x) \neq I$ or $\bar{b}(x) = -\bar{B}_x(x) \neq b(x,\xi)$ for all ξ, the algorithm does not necessarily (asymptotically) spend most time near a global minimum of $\bar{B}(\cdot)$, but the theory tells us just where it does spend most of its time. There are practical alterations of (1.1) for which $\{X_n\}$ would eventually spend most time near a global minimum, and these are discussed in Section 5. One form of the alteration requires a *specific 'cyclic' correlation among the* $\{\xi_i\}$. Very similar results hold for the diffusion

$$(1.2) \qquad dy = \bar{b}(y)dt + \frac{\sqrt{A_0}}{\sqrt{\log(t+A_1)}}\,\sigma(y)dw.$$

Section 2 defines a number of terms and results from the theory of large deviations which are useful for the further formulation of our problem. In Section 3, we treat the mean (asymptotic) escape time of $\{X_n\}$ from a set containing a stable set for the algorithm (1.1). Such quantities are important in the study of any such algorithm, since they provide useful information on the stability of the algorithm, convergences, etc. In Section 4, we treat a global problem, where there are (possibly) many stable and unstable sets for (1.1) (i.e., for the ODE $\dot{x} = Eb(x,\xi)$). The asymptotic formulas for the mean transition time between these sets is obtained, as is the (conditional) transition distribution of a certain chain associated with the asymptotic behavior - whose properties yield the mean transition and sojourn times, 'near' invariant measures, etc. Section 5 contains extensions: the form of the result for Itô equation models, the case where $-\bar{b}(x)$ is obtained from a gradient of a potential function, and applications to the problem of global minimization of a function via Monte Carlo methods, and estimates of the asymptotic measures for $\{X_n\}$ and $y(\cdot)$.

In the more standard works in 'simulated annealing' [1] to [3] the objective function values are known exactly - and the algorithm can 'move' by large steps. Here and in [8] the parameter set is not discrete and the algorithm moves only in small steps. With this constraint, the various algorithms are all quite similar - in that the transition probabilities are close. Often moving by small steps makes sense - particularly when the parameter set is not discrete. Of course, we allow sampling noise and/or imprecise measurements. The results in [8] are special cases of the results here.

Numerous variations are possible - with rather similar results, although the proofs might involve somewhat more details. E.g., the $\sigma(X_n)$ can be periodic. In the sampling noise case, this might make sense. The direction of iteration of (1.1) can be chosen at random - or several (a random number) steps can be taken in each direction. Approximations to Gaussian noise $\{\xi_n\}$ can be used - with results close to those obtained here.

2. Definitions and Assumptions

In order to simplify part of the notation later on, we actually work with slightly altered $\{\psi_n\}$. We truncate ψ_n so that $a_n|\psi_n| \to 0$ w.p.1 but the truncation level goes to ∞ as $n \to \infty$. This can be done since for each $\delta > 0$

$$P\{a_n|\psi_n| \geq \delta > 0\} \leq \exp -\delta^2/2\sigma^2 a_n^2 = \gamma_n$$

for some $\sigma^2 > 0$, where $\Sigma\gamma_n < \infty$. In calculating the action functionals below we can use either ψ_n or their truncated version, and then take limits. The result is the same. The procedure is unrestrictive, since we are interested in the 'tails' of the $\{X_n\}$ and related processes. We continue to use the ψ_n notation - but also assume that (eventually) $a_n|\psi_n|$ is as small as desired.

For the case of <u>independent and identically distributed</u> $\{\xi_n\}$, define the H, H_0 and L functionals in the usual way in the theory of large deviations: Let $\bar{b}(x) = Eb(x,\xi_n)$ and define

$$\begin{aligned} H_0(\alpha,x) &= \log E \exp \alpha'[b(x,\xi) - \bar{b}(x) + \sigma(x)\psi] \\ &= \log E \exp \alpha'[b(x,\xi) - \bar{b}(x)] + \alpha'\sigma(x)\Sigma\sigma'(x)\alpha/2. \end{aligned} \tag{2.1}$$

where $\Sigma = \text{cov}\ \psi$,

$$H(\alpha,x) = \alpha'\bar{b}(x) + H_0(\alpha,x)$$

$$L(\beta,x) = \sup_\alpha[\alpha'\beta - H(\alpha,x)] = \sup_\alpha[\alpha'(\beta - \bar{b}(x)) - H_0(\alpha,x)].$$

It is often convenient to treat $L(\cdot,\cdot)$ as a function of $\beta - \bar{b}(x)$ and x.

Generally, we assume that there is a continuous function $H_1(\alpha,x)$, differentiable in α, and a Lipschitz continuous $\bar{b}(\cdot)$, such that for any bounded stopping time ν and associated σ-algebra $B_\nu = B(\xi_{i\wedge\nu}, \psi_{i\wedge\nu}, i < \infty)$,

$$\bar{b}(x) = \lim_N \frac{1}{N} E_{B_\nu} \sum_{\nu+1}^{\nu+N} b(x,\xi_i)$$

(2.2)

$$H_1(\alpha,x) = \lim_N \frac{1}{N} \log E_{B_\nu} \exp \alpha' \sum_{n=\nu}^{\nu+N-1} [b(x,\xi_n) - \bar{b}(x)],$$

where the convergence is uniform in ω and ν and in (x,α) in any compact set; for example, any finite state ergodic Markov chain $\{\xi_n\}$ will do, as will the 'cyclic' noise of Section 5, or any sufficiently strongly (and stationary) mixing process. We now define

$$H_0(\alpha,x) = \alpha'\sigma(x)\Sigma\sigma'(x)\alpha/2 + H_1(\alpha,x)$$

and

$$H(\alpha,x) = \alpha'\bar{b}(x) + H_0(\alpha,x).$$

For each $T < \infty$, define the action functional $S_x(T,\phi)$ as equal to ∞ for ϕ not absolutely continuous, and otherwise

$$S_x(T,\phi) = \int_0^T L(\dot{\phi}(s),\phi(s))ds, \quad \phi(0) = x.$$

Let $U(x)$ denote the set $\{\beta: L(\beta,x) < \infty\}$, with closure $\bar{U}(x)$. $U(\cdot)$ is convex and upper semicontinuous (in the Hausdorff topology) in that $\lim_{x_n\to x} U(x_n) \subset U(x)$. In the i.i.d. case

$$\bar{U}(x) = \bar{b}(x) + \overline{co}[b(x,\xi) - \bar{b}(x) + \sigma(x)\psi] \equiv \bar{b}(x) + \bar{U}_0(x),$$

where $\overline{co}$ is the closed convex hull over the (a.s.) range of ξ, ψ. Information on $\bar{U}(\cdot)$ in the general case is in [6, Section 3] and in [4,5]. The $U_0(x)$ will be a set of 'control values' for the differential equation

$$\dot{x} = \bar{b}(x) + u, \quad u(t) \in U_0(x(t)). \tag{2.3}$$

$U_0(x)$ and (2.3) give approximations to the possible noise determined paths of the continuous parameter interpolation of the $\{X_n\}$, with interpolation intervals $\{a_n\}$. We always assume that $U(\cdot)$ is continuous in the Hausdorff topology.

In the i.i.d $\{\xi_i\}$ case, if there is a $\bar{\Sigma} > 0$ such that $\text{cov}[b(x,\xi) - \bar{b}(x) + \sigma(x)\psi] = \bar{\Sigma} > 0$, we say the system is <u>non-degenerate</u>. If $\bar{\Sigma}(x)$ is singular for some x, then the case is said to be <u>degenerate</u>. In general, let

$$\frac{1}{N}\,\text{cov}\sum_{1}^{N}[b(x,\xi_i) - \bar{b}(x) + \sigma(x)\psi_i] \equiv \Sigma_N(x) \xrightarrow{N} \Sigma(x)$$

$$= \begin{bmatrix} \Sigma_{11}(x) & \Sigma_{12}(x) \\ \Sigma_{21}(x) & \Sigma_{22}(x) \end{bmatrix}.$$

The system is said to be <u>non-degenerate</u> if $\Sigma(x) \geq \bar{\Sigma} > 0$. Consider the special form of the degenerate problem where $\Sigma_{11}(x) = \Sigma_{12}(x) = \Sigma_{21}(x) = 0$, $\Sigma_{22}(x) \geq \bar{\Sigma}_{22} > 0$. Then we can write (use $x = (x_1,x_2)$, $\alpha = (\alpha_1,\alpha_2)$, etc.)

$$X_{1,n+1} = X_{1,n} + a_n\bar{b}_1(X_n)$$
$$X_{2,n+1} = X_{2,n} + a_n b_2(X_n,\xi_n) + a_n\sigma(X_n)\psi_n.$$

Define

$$H_N(\alpha,x) = \frac{1}{N} \log E \exp \sum_1^N \alpha'[b(x,\xi_i) - \overline{b}(x) + \sigma(x)\psi_i],$$

and let $H_{N,\alpha}$ and $H_{N,\alpha\alpha}$ denote the gradient and Hessian matrices (with respect to α), resp. Then $H_N(\alpha,x) \geq 0$, $H_{N,\alpha}(0,x) = 0$ and $H_{N,\alpha\alpha}(0,x) = \Sigma_N(x)$. Let $\overline{K}$ denote an arbitrary compact set. Since $H_N(x,\alpha) \to H(x,\alpha) \geq 0$ and $H_N(\cdot,x)$ is convex, in the non-degenerate case $H(\cdot,x)$ is strictly convex in some neighborhood of $\alpha = 0$ which does not depend on x, for $x \in \overline{K}$. This implies that there is a $\delta_1 > 0$ such that for β in the δ_1-neighborhood $N_{\delta_1}(\overline{b}(x))$ and $x \in \overline{K}$, $L(\beta,x)$ is uniformly bounded also $L(\overline{b}(x),x) = 0$, $L_\beta(\overline{b}(x),x) = 0$ and $L(\cdot,x)$ is <u>strictly</u> convex in $N_{\delta_1}(\overline{b}(x))$. Thus $L(\beta,x) = o(|\beta-\overline{b}(x)|)$. For the above special form of the <u>degenerate</u> case, $L(\beta,x) = \infty$ unless $\beta_1 = \overline{b}_1(x)$. There is a $\delta_1 > 0$ (δ_1 not depending on x in $\overline{K}$) such that for $x \in \overline{K}$, $\beta_2 \in N_{\delta_1}(\overline{b}_2(x))$ and $\beta_1 = \overline{b}_1(x)$, $L(\beta,x)$ is uniformly bounded, also, $L_{\beta_2}(\overline{b}(x),x) = 0$ and $L(\overline{b}_1(x),\cdot,x)$ is strictly convex for $\beta_2 \in N_{\delta_1}(\overline{b}_2(x))$ and $L(\overline{b}_1(x),\beta_2,x) = o(|\beta_2-\overline{b}_2(x)|)$.

Define $\{X_n^\varepsilon\}$ and $x^\varepsilon(\cdot)$ by

$$\begin{aligned} X_{n+1}^\varepsilon &= X_n^\varepsilon + \varepsilon b(X_n^\varepsilon,\xi_n) + \varepsilon\sigma(X_n^\varepsilon)\psi_n \\ x^\varepsilon(t) &= X_n^\varepsilon \quad \text{on} \quad [n\varepsilon,n\varepsilon+\varepsilon) \end{aligned} \tag{2.5}$$

A <u>piecewise</u> <u>linear</u> interpolation could also be used to define $x^\varepsilon(\cdot)$. Under our conditions, mild alterations of the arguments in [4] can be used to show that $S_x(T,\phi)$ is an action functional for $\{x^\varepsilon(\cdot)\}$ in the sense that: for any T and any Borel set A (with interior A^0 and closure $\bar{A}$) in $C_x^r[0,T]$, the space of R^r-valued continuous functions on $[0,T]$ with initial value x,

$$
\begin{aligned}
(2.6) \qquad -\inf_{\phi\in A^0} S_x(T,\phi) &\leq \underline{\lim}_{\varepsilon}\, \varepsilon \log P_x\{x^\varepsilon(\cdot)\in A\} \\
&\leq \overline{\lim}_{\varepsilon}\, \varepsilon \log P_x\{x^\varepsilon(\cdot)\ A\} \leq -\inf_{\phi\in\bar{A}} S_x(T,\phi).
\end{aligned}
$$

For each fixed T, let $\Phi_s^x(T) = \{\phi: S_x(T,\phi) \leq s\}$, a compact set [4]. Then for any $\delta > 0$ and $d > 0$, there is an $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$ [4] ($d(\cdot,\cdot)$ denotes the appropriate distance)

$$
(2.7) \qquad P_x\{d(x^\varepsilon(\cdot),\Phi_s^x(T)) \geq \delta\} \leq \exp-(s-d)/\varepsilon.
$$

The uniformity of convergence (conditional on x,B_ν) in the definition (2.2) of $H(\alpha,x)$ has some consequences which will be quite important in the sequel. First, it is convenient to introduce some additional terminology. Let τ denote a stopping time with respect to the family of σ-algebras $\{B(\psi_i,\xi_i,\ i < t/\varepsilon)\} \equiv \{B_\varepsilon(t)\}$, and let $B_\varepsilon(\tau)$ denote the associated 'stopped' σ-algebra. Let $P_{x,B_\varepsilon(\tau)}$ denote the probability measure for $x^\varepsilon(\cdot)$, conditioned on $B_\varepsilon(\tau)$ and on $x^\varepsilon(\tau) = x$. When using this terminology, it is only necessary that the x be in the a.s. range of $x^\varepsilon(\tau)$, and we always assume this. Equivalently, we can assume that $P_{x,B_\varepsilon(\tau)}$ is the conditional distribution for the process which is reset to value x at time τ and then continues to evolve as before. Then (2.6) and (2.7) can be replaced by (2.6'), (2.7'), where the convergence in (2.6') and the bound in (2.7') are uniform in τ, ω and in x- in any compact set.

$$
\begin{aligned}
(2.6') \qquad & -\inf_{\phi\in A^0} S_x(T,\phi) \le \underline{\lim}_{\varepsilon}\ \varepsilon \log P_{x,B_\varepsilon(\tau)}\{x^\varepsilon(\tau+\cdot)\in A\} \\
& \le \overline{\lim}_{\varepsilon}\ \varepsilon \log P_{x,B_\varepsilon(\tau)}\{x^\varepsilon(\tau+\cdot)\in A\} \le -\inf_{\phi\in\bar{A}} S_x(T,\phi),
\end{aligned}
$$

$$
(2.7') \qquad P_{x,B_\varepsilon(\tau)}\{d(x^\varepsilon(\tau+\cdot),\phi_s^x(T)) \ge \delta\} \le \exp-(s-d)/\varepsilon.
$$

This result is discussed in [6] and follows from the calculations in [4]. Indeed, the convergence proofs in [4] for the unconditional form depends only on the convergence in the definition (2.2), and the estimates and convergence (2.6', 2.7') will be uniform in any parameter in which the convergence in (2.2) will be uniform.

3. The Escape Time Problem

Let K_0 denote a compact stable invariant set for

$$\dot{x} = \bar{b}(x), \tag{3.1}$$

and G a bounded open set containing K_0, with a piecewise differentiable boundary ∂G and with $\bar{G}$ in the domain of attraction of K_0; i.e., all trajectories starting in $\bar{G}$ converge to K_0. For each $n > 0$, define $\{X_k^n\}$ and $x^n(\cdot)$ by $X_0^n = x$ and

$$\begin{aligned} X_{k+1}^n &= X_k^n + a_{n+k} b(X_k^n, \xi_{n+k}) + a_{n+k}\sigma(X_k^n)\psi_{n+k} \\ x^n(t) &= X_k^n \quad \text{for} \quad t \in \left[\sum_{i=0}^{k-1} a_{n+i}, \sum_{i=0}^{k} a_{n+i}\right) \equiv [t_k^n, t_{k+1}^n) \end{aligned} \tag{3.2}$$

(3.2) is the actual process with which we work in order to study the tail of the $\{X_n\}$. Set $\tau^n = \min\{t: x^n(t) \notin G\}$. In this section, we will compute the asymptotics of $\{\tau^n\}$ for $x \in G$ under the following 'controllability' condition:

A3.1. For each $\delta > 0$ there is a ρ-neighborhood $N_\rho(K_0)$ of K_0 and $\delta_\rho > 0$, $T_\rho < \infty$, such that for each $x,y \in N_\rho(K_0)$, there is a path $\phi(\cdot)$: $\phi(0) = x$, $\phi(T_y) = y$ where $T_y \leq T_\rho$ and $S_x(T_\rho, \phi) \leq \delta$.

The condition is not very restrictive, and holds in 'typical' cases. It is a natural generalization of the usual non-degeneracy assumption which is used in large deviations work with diffusion processes - where the analog of (A3.1) always holds. For example, let $\bar{b}(x) = 0$ in K_0 and let the problem be non-degenerate (e.g., let $\sigma(x)\Sigma\sigma'(x) > 0$

on K_0). Then (A3.1) follows from $L(\beta,x) = o(|\beta-\bar{b}(x)|)$ (see above 2.5).) Alternatively, assume non-degeneracy. For each $\gamma > 0$ there is a $T_\gamma < \infty$ such that the γ-neighborhood of the path of (3.1) on $[0,T_\gamma]$, which starts at any $x \in K_0$, covers K_0. (A3.1) follows from these facts and the fact that $S(T,\phi) = 0$ for functions $\phi(\cdot)$ satisfying (3.1). In typical applications to global minimization by Monte Carlo, (A3.1) <u>and</u> (3.3) below holds, since $\operatorname{cov} \psi_n > 0$ and $\sigma(x) =$ identity matrix.

Define

$$S(x,y) = \inf_{\phi,T} \{S(t,\phi):\phi(0) = x,\phi(T) = y\}$$

and, for $x \in G$,

$$S_G(x) = \inf_{y\in\partial G} S(x,y) = \inf_{\phi,T}\{S(T,\phi):\ \phi(0) = x,\phi(T) \in \partial G\}$$

$$S_G(B) = \inf_{x\in B} S_G(x).$$

By (A3.1) and the fact that $S(T,\phi) = 0$ if $\phi(\cdot)$ satisfies (3.1), $S_G(x)$ is constant on K_0 and $S_G(x) \leq S_G(K_0)$ for $x \in G$. The $S(\cdot,\cdot)$ and $S_G(\cdot)$ are lower semicontinuous functions (this result does not require (A3.1)) [4]. Also

$$S(x,y) = \lim_{T\to\infty} \lim_{\delta\to 0} \lim_{\varepsilon\to 0} \varepsilon \log P_x\{\tau_\delta \leq T\},$$

where $\tau_\delta = \inf\{t:\ x^\varepsilon(t) \in N_\delta(y)\}$.

In Theorem 1, we will have use for the following auxiliary result.

Lemma 3.1. Under (A3.1) and the other conditions assumed above, for each $y \in G$, $S(x,y)$ (resp., $S_G(x)$) is continuous at each $x \in K_0$.

Outline of proof: The proof uses the controllability condition to construct 'nearly' optimal paths. Let $S(x,y) < \infty$. Otherwise, a similar proof yields $S(x_n,y) \to \infty$ as $x_n \to x \in K_0$. By the controllability (A3.1), for $x,x' \in K_0$, and any sequence $x_n \to x$, $S(x,x') = 0$, $S(x_n,x) \to 0$ and $S(x,x_n) \to 0$. Then the lemma follows from

$$S(x_n,y) \le S(x_n,x) + S(x,y)$$

$$S(x,y) \le S(x,x_n) + S(x_n,y),$$

$$-S(x,x_n) \le S(x_n,y) - S(x,y) \le S(x_n,x). \quad \text{Q.E.D.}$$

Let G_ρ denote a ρ-neighborhood of G.

Theorem 1. Assume (A3.1) and the other conditions above, and suppose that as $\rho \downarrow 0$

$$S_{G_\rho}(x) \downarrow S_G(x), \quad \text{some} \quad x \in K_0. \tag{3.3}$$

Then, for large enough A_0,

$$\lim_n a_n \log E_x \tau^n = S_G(K_0), \quad x \in G. \tag{3.4}$$

Remark. The continuity condition (3.3) is not very restrictive. If it doesn't hold for G, it will hold for a small perturbation of G. It always holds if $\sigma(x)E\psi\psi'\sigma(x)$ is positive definite on ∂G. Other conditions guaranteeing it, based on the 'controllability' assumption (A3.1), are in [6, Section 4]. It also holds for the particular replacement for G used in Section 4 $(R^r - \bigcup_{i\in J} N_{\mu_1}(K_i))$. The proof of Theorem 1 is an adaptation of that of Theorem 4.1 of [7].

Proof. Part 1. The asymptotic properties and result (3.4) will not change if we redefine a_n as $a_0 = a_1 = A_0$, $a_n = A_0/\log n$, for $n > 1$. Thus, after time 1, we set $A_1 = 0$. In order to obtain the lower and upper bounds on the escape times which will give (3.4), we approximate the $\{a_n\}$ by a piecewise constant process. We will actually work only with this approximation, but the general result will follow readily from it. Fix $\alpha > 1$ but close to 1. We divide the 'discrete' time into sections $\{[1,n_1), [n_1,n_2),\ldots\}$ such that the ratio of the a_n value at the start of a section (a_{n_k}) to that at the end $(a_{n_{k+1}})$ is roughly α. Thus (if n_k is not an integer, use any 'nearest' integer)

$$n_{k+1} = n_k^{\alpha}, \quad n_1 > 1. \tag{3.5}$$

For $n \in [n_k,n_{k+1})$, $k > 1$, define

$$b_n = A_0/\log n_1^{\alpha^k} \equiv \varepsilon_k = (A_0/\log n_1)/\alpha^k.$$

We call $[n_k,n_{k+1})$ the kth section. In the (continuous parameter) interpolated time scale, this section has length

$$\varepsilon_k[n_{k+1}-n_k] = \left(\frac{A_0}{\log n_1}\right)\frac{n_1^{\alpha^k}(n_1^{\alpha}-1)}{\alpha^k} = A_0A_3\exp[\alpha^k\log n_1 - k\log\alpha], \tag{3.6}$$

$$A_3 = (1/\log n_1)(n_1^{\alpha} - 1).$$

This interval is larger than $A_0\exp c_0\alpha^k$ (for some $c_0 > 0$) if k is large enough. The 'interpolated interval' $\varepsilon_k[n_{k+1}-n_k]$ is called the <u>interpolated</u> k-<u>section</u>.

We now define the analog of (3.2) with piecewise constant coefficients. Define $\{\bar{X}_n^k\}$ and $\bar{x}^k(\cdot)$ as follows. $\bar{X}_0^k = x$ and for each k

$$\bar{X}_{n+1}^k = \bar{X}_n^k + \alpha_n^k[b(\bar{X}_n^k,\xi_{n_k+n}) + \sigma(\bar{X}_n^k)\psi_{n_k+n}], \tag{3.7}$$

where $\alpha_n^k = \varepsilon_k$ for the first $(n_{k+1}-n_k)$ terms, $\alpha_n^k = \varepsilon_{k+1}$ for the next $(n_{k+2}-n_{k+1})$ terms, etc. To define the piecewise constant continuous parameter interpolation $\bar{x}^k(\cdot)$ of $\{\bar{X}_n^k\}$, we use interpolation intervals $\{\alpha_n^k\}$. Let $B_k(t)$ denote the minimal σ-algebra measuring all the data $\{\xi_i,\psi_i\}$ starting from time zero up to that used to calculate $\bar{x}^k(t)$.

Define $\tau_k = \min\{t: \bar{x}^k(\cdot) \notin G\}$. Let τ^ε denote the escape time from G for the process $x^\varepsilon(\cdot)$ introduced in (2.5). Then

$$\lim_{\varepsilon} \varepsilon \log E_x\tau^\varepsilon = S_G(K_0), \quad x \in G. \tag{3.8a}$$

((3.8a) will be obtained as a by-product of the development below.) We will show that

$$\lim_{k} \varepsilon_k \log E_x\tau_k = S_G(K_0), \quad x \in G. \tag{3.8b}$$

The theorem readily follows from this and the arbitrariness of α. The <u>key</u> to the development is the fact that replacing ε by ε_k in (3.8a) and taking limits as $k \to \infty$ yields (in the sense of logarithmic asymptotic equivalence)

$$E_x\tau^{\varepsilon_k} \sim \exp(S_G(K_0)\alpha^k/A_0),$$

and for large A_0 the ratio of the quantity in (3.6) to this expression, namely

$$N_k' \equiv \frac{(\exp c_0\alpha^k)A_0}{\exp(S_G(K_0)\alpha^k/A_0)}, \tag{3.9}$$

goes to infinity very fast as $k \to \infty$.

Part 2. Assume $S_G(K_0) < \infty$. Otherwise a similar proof yields the result. Fix $d > 0$. Choose $0 < \mu_1 < \mu_2 < \mu_3$, T_1, T_2, T_3, $(T_2 > T_3)$, $h > 0$ and $\delta > 0$ such that the following hold: (a) for any initial condition $x \in G - N_{\mu_1}(K_0)$, the path of (3.1) gets into $N_{\mu_1}(K_0)$ by time T_1 and never leaves $N_{\mu_2}(K_0)$ after that; (b) for each point $x \in \overline{N}_{\mu_2}(K_0)$ there is a path $\phi^x(\cdot)$ *such that* $\phi^x(0) = x$, $\phi^x(t) \notin G_h = N_h(G)$ (the h-neighborhood of G) for some $t \le T_3$ and

$$S_x(T_3,\phi^x) \le S_G(K_0) + d/4;$$

(c) there is a path with cost $\le d/8$ connecting any $x,y \in \overline{N}_{\mu_3}(K_0)$ in time $\le T_2 - T_3$.

The requirement (c) can be satisfied owing to (A3.1). The requirement (b) can be satisfied owing to the 'continuity' (3.3), and (a) can be satisfied since K_0 is the only limit set in $\overline{G}$ (or in G_h, for small h) for (3.1).

Part 3. Let N_k denote the number of intervals of length T in the interpolated k-section. By (3.6), there is a $c_1 > 0$ such that $N_k \ge \exp c_1\alpha^k$ for large enough A_0. We will next evaluate (3.10), an upper bound for $E_x\tau_k$.

$$E_x\tau_k \le T \sum_1^{\infty} P_x\{\tau_k > nT\} \tag{3.10}$$

Let $P_{B_k(t)}$ denote the probability measure for $\bar{x}^k(\cdot)$, conditioned on $B_k(t)$. Until after (3.14) let $nT < \varepsilon_k(n_{k+1}-n_k)$. Note that

$$P_x\{\tau_k>nT\} = E_x[1 - P_{B_k(nT-T)}\{\tau_k-(nT-T)\le T\}]I_{\{\tau_k>nT-T\}}. \tag{3.11}$$

We have (A and B are defined by the two events described in the middle term of (3.12)) for $T = T_1 + T_2$

$$\begin{aligned} &P_{B_k(nT-T)}\{\tau_k-(nT-T) \le T_1 + T_2\}I_{\{\tau_k>nT-T\}} \\ &\ge P_{B_k(nT-T)}\{\bar{x}^k(\cdot) \text{ goes to } N_{\mu_1}(K_0) \text{ and then stays in } N_{\mu_2}(K_0), \\ &\qquad \text{all on } [nT-T,nT-T+T_1], \text{ then leaves } G \text{ on } [nT-T+T_1,nT]\} \\ &\qquad \text{starting in } \bar{N}_{\mu_2}(K_0)\}I_{\{\tau_k>nT-T\}} \\ &= P_{B_k(nT-T)}\{A \cap B\}I_{\{\tau_k>nT-T\}}. \end{aligned} \tag{3.12}$$

We have

$$\inf_{\omega:\bar{x}^k(nT-T)\in G} P_{B_k(nT-T)}\{A\} \ge \frac{1}{2}$$

for all ω for large k. This follows from (2.7') together with the fact that $S_y(T,\phi) = 0$ if and only if $\phi(\cdot)$ satisfies (3.1), and the fact that all trajectories of (3.1) starting in $\bar{G}$ stay in $N_{\mu_2}(K_0)$ after time T_1. Then, for $y \in N_{\mu_2}(K_0)$ and large k, (2.6') yields

$$P_{B_k(nT-T+T_1)}\{B\} I_{\{\tau_k > nT-T\}}$$

(3.13)
$$\geq \inf_{y,\omega:\bar{x}^k(nT-T+T_1)=y\in N_{\mu_2}(K_0)} P_{B_k(nT-T+T_1)} \cdot \{\sup_{0\leq t\leq T_3} |\phi^y(t)-\bar{x}^k(nT-T+T_1+t)| < h/4\} I_{\{\tau_k>nT-T\}}$$

$$\geq [\exp - [S_G(K_0) + d/2]/\varepsilon_k] I_{\{\tau_k>nT-T\}}$$

Combining the above estimates yields that (3.11) is bounded above by

(3.14)
$$P_x\{\tau_k > nT\} \leq E_x[(1 - \frac{1}{2}\exp-(S_G(K_0) + d/2)/\varepsilon_k] I_{\{\tau_k>nT-T\}}.$$

Define

$$\beta_k = [1 - \exp - (S_G(K_0) + d/2)/\varepsilon_k]^{N_k}.$$

Now iterate (3.14) up to n: $nT = \varepsilon_k(n_{k+1}-n_k)$, then use ε_{k+1} for the next N_{k+1}, etc. Doing this and substituting into (3.10) yields that for large k,

$$\begin{aligned} E\tau_k &\leq T \sum_{n=0}^{N_k} [1 - \exp - (S_G(K_0) + d/?)/\varepsilon_k]^n \\ &+ T\,\beta_k \sum_{n=0}^{N_{k+1}} [1 - \exp - (S_G(K_0) + d/2)/\varepsilon_{k+1}]^n \\ &+ T\,\beta_k\beta_{k+1} \sum_{n=0}^{N_{k+2}} [1 - \exp - (S_G(K_0) + d/2)/\varepsilon_{k+2}]^n + \cdots . \end{aligned}$$

In order to estimate the terms in the sum beyond the first, note that

$$\beta_m \exp[S_G(K_0) + d/2]/\varepsilon_{m+1}$$

$$\leq \exp-[N_m \exp-(S_G(K_0) + d/2)/\varepsilon_m] \exp[S_G(K_0) + d/2]/\varepsilon_{m+1},$$

which (since $N_m \geq \exp c_1 \alpha^m$ for some $c_1 > 0$ if m and A_0 are large) is a term of a summable series. Thus for large k and A_0,

$$E_x\tau_k \leq T \sum_0^\infty [1 - \exp-(S_G(K_0) + d/2)/\varepsilon_k]^n + \text{constant} \tag{3.15}$$

$$\leq \exp(S_G(K_0) + d)/\varepsilon_k.$$

Part 4. For the rest of the proof we let $\{\xi_n\}$ be mutually independent. This is for notational convenience only. It allows us to avoid the notation associated with the conditioning used in (e.g.) (3.12) and (3.13). In general, we work as in the last part by using appropriate conditioning and taking $\sup_\omega$ or $\inf_\omega$, as appropriate. In fact, the proof of Theorem 2 in Section 4 uses the 'full conditioning' argument. We will need the following lemma, whose proof is only a slight modification of that of Lemma 2.2 of [7, Chapter 4] or Lemma 1.9 of [7, Chapter 6] and is omitted. The second part of the lemma will be used in Theorem 2 below (it does not assume mutual independence of $\{\xi_i\}$).

Lemma 2. For each small $\alpha > 0$, there are $c > 0$, $T_0 < \infty$, $\varepsilon_0 > 0$, such that for $\varepsilon \leq \varepsilon_0$ and all $y \in \bar{G} - N_\alpha(K_0)$ and all T_4,

$$P_y\{\tau_\alpha^\varepsilon > T_4\} \leq \exp - c(T_4 - T_0)/\varepsilon$$

where

$$\tau_\alpha^\varepsilon = \inf\{t: x^\varepsilon(t) \notin G - N_\alpha(K_0)\}.$$

More generally, let τ denote a stopping time and let K be compact but not contain an entire limit set for (3.1). Define $\tau_K^\varepsilon = \inf\{t: x^\varepsilon(\tau+t) \notin K\}$. Then

$$P_{y,B_\varepsilon(\tau)}\{\tau_K^\varepsilon - \tau > T_4\} \leq \exp - c(T_4 - T_0)/\varepsilon,$$

for all finite stopping times τ and $y \in K$ and $\varepsilon \leq \varepsilon_0$. ($P_{y,B_\varepsilon(\tau)}$ was defined above (2.6').)

We now proceed very similarly to [7, p. 125-6]. Since

$$\sum_{n=N_k}^{\infty} [1 - \exp - (S_G(K_0) + d/2)/\varepsilon_k]^n \xrightarrow{k} 0,$$

and the contribution to the mean value of τ_k from paths which do not exit G before ε_{k+1} is used is vanishingly small as $k \to \infty$, in calculating a lower bound on the escape time we can and will assume for (3.7) that $\alpha_n^k \equiv \varepsilon_k$ for all n (or, equivalently, work with $x^\varepsilon(\cdot)$, for $\varepsilon = \varepsilon_k$). Define $F_1 = \bar{N}_{\mu_3}(K_0) - N_{\mu_2}(K_0)$, $F_2 = \bar{N}_{\mu_1}(K_0) \cup (R^r - G)$ and define the stopping times $\{\sigma_i, \rho_i\}$ by $\rho_0 = 0$

$$\sigma_i = \inf\{t > \rho_i : \bar{x}^k(t) \in F_1\},$$

$$\rho_i = \inf\{t > \sigma_{i-1} : \bar{x}^k(t) \in F_2\}.$$

The only way for $\bar{x}^k(\cdot)$ to jump from the exterior of $\bar{N}_{\mu_3}(K_0)$ into $N_{\mu_2}(K_0)$ is if it is pushed there by a very large value of ψ_n. But this is ruled out by the comments made in the beginning of Section 2.

For $x \in \bar{N}_{\mu_1}(K_0)$ and any $T_4 < \infty$,

$$P_x\{\bar{x}^k(\rho_1) \in (R^r - G)\} \leq \max_{y \in F_1} P_y\{\tau_k = \rho_1 \leq T_4\} \tag{3.16}$$
$$+ P_x\{\tau_k = \rho_1 > T_4\}$$

By Lemma 2, for each $M < \infty$ there is a $T_4 < \infty$ such that the far right hand term of (3.16) is less than $\exp -M/\varepsilon_k$ for large k.

Recall that we chose the μ_i such that $S(x',y') \leq d/4$ for $x',y' \in \bar{N}_{\mu_3}(K_0)$. By the compactness (for each s,y) and upper semicontinuity (in s,y) of the sets $\Phi_s^y(t)$, there is a $\delta_1 > 0$ (not depending on y) such that the paths which start at $y \in \bar{N}_{\mu_3}(K_0)$ and exit G before T_4 are at a distance of at least δ_1 from the set $\Phi^y_{S_G(K_0)-d/2}(T_4)$. (The minimum value of $S_y(T_4,\phi)$ for such an exiting path must be at least $S_G(K_0) - d/4$.) Then it follows from (2.7) that for $y \in F_1$

$$P_y\{\tau_k < T_4\} \leq \exp -[S_G(K_0)-d]/\varepsilon_k, \tag{3.17}$$

for large k. In fact by the just cited uppersemicontinuity and compactness, we can write $\max_{y \in F_1} P_y\{\tau_k < T_4\}$ in (3.17). Then, for a large fixed M and all large k,

$$P_x\{\bar{x}^k(\rho_1) \notin G\} \leq \exp -(S_G(K_0)-2d)/\varepsilon_k, \quad x \in \bar{N}_{\mu_1}(K_0).$$

Define $\nu = \min\{n: \bar{x}^k(\rho_n) \notin G\}$. Then for $x \in \bar{N}_{\mu_1}(K_0))$,

$$P_x\{\nu>n\} = P_x\{\vec{x}^k(\rho_i) \in \overline{N}_{\mu_1}(K_0),\ \text{all}\ i \le n\}$$

$$= E_x P_x\{\vec{x}^k(\rho_n) \in \overline{N}_{\mu_1}(K_0) | \vec{x}^k(\rho_{n-1})\} I_{\{\nu>n-1\}}$$

$$\ge \inf_{y \in \overline{N}_{\mu_1}(K_0)} P_y\{\vec{x}^k(\rho_1) \in \overline{N}_{\mu_1}(K_0)\} P_x\{\nu>n-1\}$$

$$\ge (1 - \exp - (S_G(K_0) - 2d)/\varepsilon_k)^n.$$

For each $\{\mu_i\}$, there is a $t_1 > 0$ such that $\inf_y E_y(\rho_1 - \sigma_0) \ge t_1$. Thus

$$E_x\tau_k = \sum_1^\infty E_x I_{\{\nu \ge n\}}(\rho_n - \rho_{n-1})$$

$$\ge \sum_1^\infty E_x I_{\{\nu \ge n\}}(\rho_n - \sigma_{n-1})$$

$$\ge \sum_1^\infty P_x\{\nu \ge n\} \inf_y E_y(\rho_1 - \sigma_0)$$

$$\ge (\text{constant}) \exp(S_G(K_0) - 2d)/\varepsilon_k.$$

This, (3.15), and the arbitrariness of d yield (3.8b). Q.E.D.

<u>Remark</u>. The proof with use of coefficients $\alpha_n^k = a_{n+k}$ follows readily from the above proof and the fact that we can choose $\alpha > 1$ arbitrarily close to 1. For all practical purposes, the 'piecewise constant' α_n^k can be used in lieu of the a_{n+k}.

4. Asymptotic (Large Time) Properties of $\{X_n^k\}$

In this section, we obtain results analogous to those in [7, Chapter 6] for the $\{X_m^n\}$ and $\{x^n(\cdot)\}$ of (3.2). Again, we use the 'intermediate' processes $\{\bar{x}^k(\cdot)\}$ with piecewise constant coefficients to obtain the results. Let $I = \{1,\ldots,m\}$ and let $K_1,\ldots,K_m$ denote a collection of disjoint compact sets, each of which is an invariant set for (3.1), and such that $\bigcup_i K_i$ contains all the limit sets for (3.1). If $S(x,y) = 0$ for all x,y in any set K, let that K be one of the K_i. The collection $\{K_i\}$ contains all the stable (and unstable) sets for the algorithms (1.1), (2.5) and (3.2), and it is of interest to study the asymptotic statistics of the movement from a neighborhood of one of the K_i to a neighborhood of another. This is particularly important for an understanding of the use of (1.1) for global minimization (or 'near' minimization) by Monte Carlo.

We make some additional assumptions.

A4.1. The controllability assumption (A3.1) holds for each K_i, $i = 1,\ldots,m$ replacing K_0 there.

Define $S_{ij} = S(K_i,K_j) = \inf_{x\in K_i, y\in K_j} S(x,y)$. By (A4.1), $S_{ij} = S(x,y)$ for any $x \in K_i$ and $y \in K_j$ and $S(x,y) = 0$ for $x,y \in K_i$. Also, by an argument like that of Lemma 1, $S(x,y)$ *is continuous in* x,y, *for* $x,y \in \bigcup_i K_i$.

It is useful to be able to bound the paths $x^\varepsilon(\cdot)$, $\bar{x}^n(\cdot)$, etc. There are several ways of doing this. Perhaps the simplest is to project them back onto some (large) set D_1 - if they ever leave D_1. This idea involves a number of new considerations and details. A reasonable alternative is to fix the dynamics such that for some compact set (a sphere, for example) D_1, all paths remain in D_1. This is not a restriction in applications, since in the simulations we can always add a penalty function and choose $\sigma(x)$, or otherwise fix the dynamics for large $|x|$ to guarantee bounded paths. For simplicity assume

A4.2. There is a sphere D_1 such that D_1 contains $\bigcup_i K_i$ in its interior and $\sigma(x) \to 0$ as $x \to \partial D_1$ and the trajectories of $x^\varepsilon(\cdot)$, $x^n(\cdot)$ stay in D_1. All paths of (3.1) starting in D_1 stay in D_1.

By (A4.2), we can assume that for small $\delta > 0$, any δ-optimal path connecting a small neighborhood of K_i with a small neighborhood of K_j does not leave D_1. I.e., we can assume that for small $\delta > 0$, if $\phi(\cdot)$ is such that $\phi(0) = x$, $\phi(T) = y$, $x \in$ small neighborhood of K_i, $y \in$ *small neighborhood* of K_j and $S_x(T,\phi) \le S_{ij} + \delta$, then $\phi(t) \in D_1$, $t \le T$.

Let μ_i be defined as in Theorem 1 but add (fixed henceforth) a $\mu_4 > \mu_3$, with the $\{\overline{N}_{\mu_4}(K_i)\}$ disjoint; define $g_i = \overline{N}_{\mu_1}(K_i)$ and $\Gamma_i = \overline{N}_{\mu_3}(K_i) - N_{\mu_2}(K_i)$. The natural analog of the scheme in [7, Chapter 6] for getting the asymptotics of $\{x^n(\cdot)\}$ or $\{\overline{x}^k(\cdot)\}$ involves estimating the probability of the process going from g_i to Γ_i and then to g_j, $j \ne i$, and then calculating the mean times via the particular formulas developed in [7] involving products of the probabilities of various chains connecting the $\{\Gamma_i, g_j\}$. With a few modifications, the results carry over to our case. We first reproduce some of the notation in [7], adapted to our case. The proofs here will be simpler than these in [7], since the set D in [7, Chapter 6] is replaced here by a set of the form $R^r - \bigcup_{j\in J} g_j$, for some subset $J \subset I$, and the $N_{\mu_i}(K_j)$ are 'small' neighborhoods.

Let J denote a subset of I with ℓ members where $\ell < m$. Define g_J by $g_J = \bigcup_{i\in J} g_i$. By slightly altering the $N_{\mu_i}(K_j)$ we can assume that the boundaries are as smooth as desired. A J-<u>graph</u> is defined to be a set of $m-\ell$ arrows $\{\gamma\to\delta\}$ connecting points in I, where $\gamma \in I-J$, $\delta \in I$ and there are no cycles, and each point in $I-J$ has one and only one arrow leaving it. $G(J)$ denotes the collection of J-graphs. By the symbol $g \in G(i \not\to J)$, we mean a collection of $m-\ell-1$ arrows $\{\gamma\to\delta\}$, without cycles, where $i \in I-J$, $\gamma \in I-J$, $\delta \in I$ and <u>not</u> containing chains leading from i to J.

We also use the following definitions. <u>Note that our</u> S <u>in the</u> V <u>of</u> [7]. Again, the notation is adapted from [7, Chapter 6]. Define

$$\tilde{S}_{ij} = \inf\{S(T,\phi) : \phi(0) \in K_i,\ \phi(T) \in K_j,\ \phi(t) \notin \bigcup_{s\neq i,j} K_s,\ t < T\}$$

$$= \infty \text{ if the above set is empty}$$

$$W_J = \min_{g\in G(J)} \sum_{(\gamma\to\delta)\in g} S_{\gamma\delta}$$

(4.1)

$$M_J(K_i) = \min_{g\in G(i\not\to J)} \sum_{(\gamma\to\delta)\in g} S_{\gamma\delta}.$$

Let τ_J^n, $\tau_{k,J}$, and τ_J^ε denote the first entrance times into the set g_J for the $x^n(\cdot)$, $\bar{x}^k(\cdot)$ and $x^\varepsilon(\cdot)$ processes, respectively.

Theorem 2. Under our conditions, for large A_0,

$$(4.2)\qquad \begin{aligned}\lim_n a_n \log E_x\tau_J^n = \lim_k \varepsilon_k E_x \log \tau_{k,J} = \lim_\varepsilon \varepsilon \log E_x\tau_J^\varepsilon \\ = W_j - M_J(K_i),\end{aligned}$$

uniformly for x in any small enough neighborhood of any K_i.

Remark. If x is not very close to some K_i, then the path tends to a small neighborhood of some K_i 'very fast'. This fact and the theorem are enough to give us the relative asymptotic times that $\{X_n\}$ spends in any set.

Proof: If the set G in Theorem 1 is replaced by $R^r - g_J = \bigcup_J g_i$, then the 'continuity' condition (3.3) is not needed owing to (A4.2), which allows trajectories hitting $\partial G = \partial g_J$ to be extended into the interior of g_J at 'small' extra cost, if μ_1 is small. The above G also corresponds to the set D in [7, Chapter 6].

We will prove only the second two equalities of (4.2) for arbitrary $\alpha > 1$ and under the condition that all $S_{ij} = S(K_i,K_j) < \infty$. The equivalence of the first two terms follows from the calculations below. We proceed as follows. First show the middle equality in (4.2), then work with $x^\varepsilon(\cdot)$. The proof requires Lemma 3 below (our analog of Lemma 2.1 of [7, Chapter 6]). With this lemma the proof can be readily completed.

The proof of the second equality in (4.1) is similar to that in Theorem 1. Fix $d > 0$. Choose T_1 such that the paths of (3.1) starting anywhere in D_1 get to $\cup g_j$ by time T_1 and then stay there (for the appropriate small μ_i). There is such a T_1 since all the limit sets of (3.1) are strictly inside $\bigcup_i g_i$. There are $T_2 < \infty$ and $\phi^{ij}(\cdot) \in C_x[0,T_2]$, $x \in K_i$, $i \in I-J$, $j \in I$, such that $\phi^{ij}(t) \in N_{\mu_1/2}(K_j)$ for some $t \leq T_2$ and

$$S_x(T_2,\phi^{ij}) \leq S_{ij} + d/4,$$

$$S(x,x') \leq d/4 \quad \text{for} \quad (x,x') \in \bar{N}_{\mu_3}(K_i) \quad \text{and each} \quad i.$$

We can also suppose (see proof of Theorem 1) that

$$P_{B_\varepsilon(\tau),x}\{x^\varepsilon(\tau+T_1) \in \bigcup_i g_i\} \geq \frac{1}{2}, \quad x \in D_1.$$

Set $T = T_1 + T_2$ as in Theorem 1, and <u>define</u> $S_0 = \max_{i,j}[S_{ij},\tilde{S}_{ij}]+d$. Using an argument analogous to that in Theorem 1 (part 3) yields that the contribution of the time that α_n^k equals ε_m (for $m > k$) to the mean hitting time $E_x\tau_{k,J}$ is bounded above by the expression

$$\prod_1^{N_m} (1 - \exp{-Q_i^m/\varepsilon_m}) \sum_0^\infty \prod_0^n (1 - \exp{-Q_i^{m+1}/\varepsilon_{m+1}}),$$

where the Q_i^m satisfy $Q_i^m \leq S_0/A_0$. Since $N_m \geq \exp c_1\alpha^m$ for some $c_1 > 0$ and large m (N_m is defined at the start of part 3 of proof of Theorem 1) the above quantity is bounded by

$$\exp[-(\exp c_1\alpha^m)\exp(-S_0\alpha^m/A_0)]\exp S_0\alpha^{m+1}/A_0$$

which (for large A_0) goes to zero faster than a geometric series as $m \to \infty$. Also, as in Theorem 1, the contribution to the mean hitting time ($E_x\tau_{k,J}$) of the part of the path beyond the first N_k interpolated T-intervals is asymptotically (as $k \to \infty$) negligible. Thus the first equality of (4.2) holds and we need only work with $x^{\varepsilon}(\cdot)$.

Define τ_n, σ_n by $\tau_0 = 0$ and

$$\sigma_n = \inf\{t > \tau_n : x^{\varepsilon}(\cdot) \in \bigcup_i \Gamma_i\}$$

$$\tau_n = \inf\{t > \sigma_{n-1} : x^{\varepsilon}(\cdot) \in \bigcup_i g_i\}$$

Let $Z_n = x^{\varepsilon}(\tau_n)$. In the following lemma, the 'conditional transition' probabilities for $\{Z_n\}$ will be estimated.

Lemma 3. Fix $d > 0$. There is an $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$ and all i,j and $x \in \Gamma_i$ and $n \geq 1$,

$$\exp-(\tilde{S}_{ij} + d/4)/\varepsilon \leq P_{x,B_{\varepsilon}(\tau_n)}\{Z_{n+1} \in g_j\} \leq \exp-(\tilde{S}_{ij} - d/4)/\varepsilon. \tag{4.3}$$

($P_{x,B_{\varepsilon}(\tau)}$ is defined above (2.6').)

Proof: Fix i,j, $i \neq j$, and $d > 0$. There are small $\{\mu_i\}$, $\mu_i > 0$, $\delta_0 > 0$ and $t_1 < \infty$ such that: for each $x \in g_i$, there is a path $\phi^x_{ij}(\cdot)$ on $[0,t_1]$ connecting x to K_i, then K_i to K_j and after leaving $N_{\mu_2}(K_i)$ the distance of the path from g_i and from g_s $(s \neq j)$ is $\geq \delta_0$, and for which $S_x(t_1,\phi^x_{ij}) \leq \tilde{S}_{ij} + d/4$. There is an $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$ and $\delta_1 = \frac{1}{2}\min(\delta_0,\mu_1)$ and $x^{\varepsilon}(\tau_n) = x \in \Gamma_i$ we have

$$\begin{aligned} P_{x,B_{\varepsilon}(\tau_n)}\{Z_{n+1} \in g_j\} &\geq P_{x,B_{\varepsilon}(\tau_n)}\{\sup_{0\leq t\leq t_1} |x^{\varepsilon}(\tau_n+t)-\phi^x_{ij}(t)| \leq \delta_1\} \\ &\geq \exp-(\tilde{S}_{ij} + d/2)/\varepsilon. \end{aligned} \tag{4.4}$$

To get the reverse inequality, note that for any $t_2 < \infty$

$$P_{x,B_\varepsilon(\tau_n)}\{Z_{n+1} \in g_j\} \le \sup_{y\in\Gamma_i,\omega} P_{y,B_\varepsilon(\sigma_n)}\{Z_{n+1} \in g_j\}$$

(4.5)
$$\le \sup_{y\in\Gamma_i,\omega} P_{y,B_\varepsilon(\sigma_n)}\{\tau_{n+1} \ge t_2\}$$

$$+ \sup_{y\in\Gamma_i,\omega} P_{y,B_\varepsilon(\sigma_n)}\{\tau_{n+1} \le t_2,\ Z_{n+1} \in g_j\}.$$

By Lemma 2, for any $M < \infty$ there is a $t_2 < \infty$ such that for small ε the first term on the right is $\le \exp -M/\varepsilon$.

If $x^\varepsilon(\sigma_n) = x \in \Gamma_i$ and $x^\varepsilon(\sigma_n+t') \in g_j$ for some $t' \le t_2$ and if $x^\varepsilon(\sigma_n+t) \notin K_s$, $s \ne i,j$ for $t \le t'$, then for small $\{\mu_j\}$ there is a $\delta_2 > 0$ such that

(4.6)
$$\sup_{\substack{0\le t\le t_2 \\ \phi\in\Phi^x_{\tilde{S}_{ij}-d/2}(t_2)}} |x^\varepsilon(\sigma_n+t) - \phi(t)| \ge \delta_2.$$

But for small ε and $\{\mu_j\}$, (2.7') implies that

(4.7)
$$P_{x,B_\varepsilon(\sigma_n)}\{\text{event defined by (4.6)}\} \le \exp - (\tilde{S}_{ij}-d)/\varepsilon$$

for all $x \in \Gamma_i$ and ω (see also the related argument below (3.16)). The far right hand term of (4.5) is bounded above by (4.7). <u>This completes the proof of the lemma since</u> M <u>is arbitrary</u>.

We now return to the proof of the theorem. Let $\{Z_n^J\}$ denote the $\{Z_n\}$ process <u>stopped</u> on firts reaching g_j. We have

(4.8)
$$E_x\tau_J^\varepsilon = \sum_0^\infty E_x I_{\{Z_n^J \notin g_J\}} E_{Z_n,B_\varepsilon(\tau_n)}(\tau_{n+1}-\tau_n)$$

For any $d > 0$, the argument of Theorem 1 yields that (for small fixed $\{\mu_j\}$) there is a $t_0 > 0$ (depending perhaps on $\{\mu_j\}$) such that for small ε, $t_0 \leq E_{x,B_\varepsilon(\tau_n)}(\tau_{n+1}-\tau_n) \leq \exp d/\varepsilon$, $x \in \Gamma_i$ (to get the r.h.s. just let G decrease to a small neighborhood of K_0 in Theorem 1). Thus, it is enough to estimate (4.8) without the $(\tau_{n+1}-\tau_n)$ component. In [7, Theorem 5.3 of Chapter 6], estimates which are the equivalent to those of Lemma 3 for the problem in [7] (those of Lemma 2.1 of Chapter 6 there) are used to show that

$$\lim_{\varepsilon} \varepsilon \log \sum_{0}^{\infty} E_x I_{\{Z_n^J \notin g_J\}} = W_J - M_J(K_i),$$

for x in a small neighborhood of K_i. Q.E.D.

5. Extensions and Comments

Cycling and asymptotic movement among the sets K_i. Let $J = \{j\}$, and fix i, where K_i and K_j are stable. The unstable sets K_ℓ are 'transient', in the sense that if K_ℓ is not stable then there is a $j \neq \ell$ such that $S_{\ell j} = 0$. Let $S_{ij} < S_{ik}$ for all $k \neq i,j$. Then Theorem 2 implies that

$$\lim_n a_n E_x \log \tau_J^n = S_{ij}, \quad x \in \text{small neighborhood of } K_i,$$

and with a 'very high' probability, K_j will be the successor state to K_i. This is almost obvious, since if (e.g.) the optimal graph in the calculation of W_J involves a link $i \to k \neq j$, then cutting that link and replacing it by the $i \to j$ link further reduces the value of W_J. The M_J is treated similarly. As in [7, Chapter 6], the asymptotic behavior can be described via 'cycles'. There will be groups of the K_i such that for a long time the process will cycle between states within a group, then switch to another group and cycle between its states. At the next higher level, there will be a cycling between these groups. The groups themselves can be formed into higher order groups, and cycling between these described, etc. The notation is involved, but the procedure to get the mean times for the transitions within any order of the hierarchy is quite similar to that in [7, Chapter 6, Section 6], and is based only on the analog of Theorem 2 and Lemma 3 for the problem in [7]. The procedure yields the (asymptotic) mean time spent in the various states.

Itô equations. Let

$$\text{(5.1)} \qquad dy = \bar{b}(y)dt + a(t)\sigma(y)dw, \quad a^2(t) = A_0/\log(t+A_1),$$

where $w(\cdot)$ is a standard R^r-valued Wiener process. Define $y^n(\cdot)$ by

$$\text{(5.2)} \qquad dy^n = \bar{b}y^n\, dt + a(n+t)\sigma(y^n)dw, \quad t \geq 0.$$

If $\sigma(y)\sigma'(y)$ is positive definite in the interior of D_1 (see (A4.2)), then the action functional for $\{y^n(\cdot)\}$ is

$$S_x(T,\phi) = \frac{1}{2}\int_0^T (\dot{\phi}(s)-\bar{b}(\phi(s)))'[\sigma(\phi(s))\sigma'(\phi(s))]^{-1}(\dot{\phi}(s)-\bar{b}(\phi(s)))ds$$

for $\phi(\cdot)$ absolutely continuous, and it equals infinity otherwise. In general

$$S_x(T,\phi) = \int_0^T L(\dot{\phi}(s),\phi(s))ds$$

where

$$L(\beta,x) = \sup_\alpha[\alpha'(\beta-\bar{b}(x)) - \frac{1}{2}\alpha'\sigma(x)\sigma'(x)\alpha].$$

The obvious analogs of Theorems 1, 2 and Lemma 3 hold with a_n replaced by $a^2(n)$ and ε by ε^2.

Invariant measures for the $y(\cdot)$ of (1.2), (5.1). Let $\sigma(x)\sigma'(x)$ be bounded and uniformly positive definite in the interior of D_1 and assume that there are only a finite number of K_i and let all trajectories of (3.1) starting in D_1 stay in D_1. Let $y^\varepsilon(\cdot)$ denote the solution to (1.2) with

$a(t)$ replaced by ε. In [7, Chapter 6, Theorem 4.1 to 4.3] an expression for the invariant measure v_ε of $y^\varepsilon(\cdot)$ is given (for small $\varepsilon > 0$). Let $v(t)$ denote the measure of $y(t)$. Then $v(t) - v_{a(t)} \to$ zero measure weakly. Thus, for large t, the measure of $y(t)$ is very close to that of the stationary measure of $y^\varepsilon(\cdot)$ for $\varepsilon = a(t)$. We will not go through the details, but they follow from the following considerations. Replace $a(n+t)$ by a piecewise constant approximation as in Theorem 1; i.e., use $\vec{y}^k(\cdot)$, where we define (for any $\alpha > 1$ and some $T_1 > 1$) for each k

(5.3) $$d\vec{y}^k = \bar{b}(\vec{y}^k)dt + a(k,t)\sigma(\vec{y}^k)dw,$$

$$T_{n+1} = T_n^\alpha = T_1^{\alpha^n},\ n \geq 1,\ \varepsilon_k^2 = A_0/\alpha^k,\quad T_0 = 0$$

$$\alpha(k,t) = \varepsilon_k \quad \text{on} \quad [0, T_{k+1}-T_k)$$

$$= \varepsilon_{k+1} \quad \text{on} \quad [T_{k+1}-T_k, T_{k+2}-T_k), \text{ etc.}$$

The measure v_ε in [7] is obtained from the invariant measure of the $\{Z_n\}$ process, where we define $\{Z_n\}$ here as in Theorem 2, but using the $\vec{y}^k(\cdot)$ of (5.3) instead of the $\{\vec{x}^k(\cdot)\}$. In fact if v_ε^Z denotes the invariant measure of the $\{Z_n\}$ for parameter ε, and $g = \cup g_i$, then [7, (4.1) of Chapter 6]

$$v_\varepsilon(B) = \int_g v_\varepsilon^Z(dy)E_y \int_0^{\tau_1} I_B(y^\varepsilon(t))dt.$$

There is an $\bar{A}_0 < \infty$ such that for $A_0 \geq \bar{A}_0$, the number of transitions of $\{Z_n\}$ on the $[T_k, T_{k+1})$ interval increases rapidly enough as $k \to \infty$ so that a 'near' steady state is reached before the end of the k^{th}-interval, for large k. To see this, note the following: (a) all $\tilde{S}_{ij} < \infty$ and $S_{ij} < \infty$; (b) for any $d > 0$, the maximum modulus of the eigenvalues (with modulus less than unity) of the transition probabilities of the chain $\{Z_n\}$ on the k^{th}-interval is* $\leq 1 - \exp-(S_0+d)/a^2(k) = 1 - \exp - (S_0+d)\alpha^k/A_0$ for large k; (c) the length of the k^{th} interval is $\geq \exp c_1\alpha^k$ for some $c_1 > 0$. Now, let $M_k = \exp c_2\alpha^k$ for $0 < c_2 < c_1$. Then

$$P\{\sum_1^{M_k} (\tau_{n+1}-\tau_n) \leq \exp c_1\alpha^k\}$$

$$\leq \frac{M_k \exp[S_0+d]\alpha^k/A_0}{\exp c_1\alpha^k} \leq \exp - c_3\alpha^k$$

for some $c_3 > 0$ for large enough A_0. Finally, note that

$$[1 - \exp - (S_0+d)\alpha^k/A_0]^{M_k} \leq \exp - c_4\alpha^k,$$

for some $c_4 > 0$ *if* A_0 *is large enough.* The assertion concerning convergence to the invariant measure follows from this.

<u>The potential case.</u> <u>Let</u> $\bar{b}(x) = -\bar{B}_x(x)$ <u>and use the process</u> $y(\cdot)$ <u>of</u> (1.2). For simplicity, add a penalty function so that $\bar{b}(x)$ points strictly inward on ∂D_1 for some sphere D_1. Let $\bar{B}(\cdot)$ be continuously differentiable and assume that there are only a finite number of the compact K_i introduced in Section 4, and that $\bar{B}(x) \to \infty$ as $|x| \to \infty$. Let $\sigma(x) = I$ except close to ∂D_1. Since $\sigma(x)\sigma'(x)$ = identity matrix 'inside' D_1, (A4.1) holds.

For this case and x,y not close to D_1, $S(x,t)$ has a simple characterization as

$$\inf_{\phi}\{\text{sums of (positive) } \underline{\text{increases}} \text{ in } \bar{B}(\phi(\cdot)) \text{ as } \phi(t) \text{ moves from } x \text{ to } y\},$$

---

* Recall $S_0 \equiv \max_{i,j}[S_{ij},\hat{S}_{ij}] + d$

where the inf is over all differentiable paths connecting x and y. There is a similar definition for $\tilde{S}_{ij}$ and S_{ij}.

The same comments apply to the system (1.1) if $b(x,\xi) = \bar{b}(x)$. In these cases the invariant measure v_ε is concentrated on an arbitrarily small neighborhood of the set of global minima of $\bar{B}(\cdot)$ for small ε [7]. Let v^n denote the measure of X_n. Then we have that $v(t)$ and v^n are both ultimately concentrated near the set of global minima of $\bar{B}(\cdot)$ also. This includes the 'annealing' result of [8].

Global functiom minimization via Monte Carlo. In many applications*, one can choose the noise ξ_n in $b(X_n,\xi_n)$, and often there are choices which greatly enhance the search. Let $\bar{b}(x) = -\bar{B}_x(x)$, where $\bar{B}(\cdot)$ and $\sigma(\cdot)$ satisfy the conditions in the above 'potential case' subsection. Then both (A4.1) and (A4.2) hold. For each m, choose $\bar{\xi}_i^m$, $i \le m$, such that** $b(x,\xi) = -B_x(x,\xi)$, $\bar{B}(x) = EB(x,\xi)$ and also such that

$$\frac{1}{m}\sum_1^m b(x,\bar{\xi}_i^m) \equiv \bar{b}^m(x) \to \bar{b}(x)$$

uniformly for x in any compact set. Define $\xi_{km+i} = \bar{\xi}_i^m$ for $k = 0,1,\ldots$. We use

$$X_{n+1} = X_n + a_n b(X_n,\xi_n) + a_n\psi_n \tag{5.4}$$

and the $x^n(\cdot)$, $\bar{x}^k(\cdot)$, $x^\varepsilon(\cdot)$ obtained from it, as in the previous sections. With this scheme, the measure of X_n will ultimately be concentrated near the set of global minimia of $\frac{1}{m}\sum_1^m B(\cdot,\xi_i)$.

Let $S_x^m(T,\phi)$ be the action functional which corresponds to $x^n(\cdot)$ for given m. Then

---

* In Monte Carlo optimization by simulation.

**We observe the noise corrupted function $B(x,\xi)$ and its gradient, where $EB(x,\xi) = \bar{B}(x)$.

$$S_x^m(T,\phi) = \int_0^T L^m(\dot{\phi}(s),\phi(s))ds,$$

where

$$L^m(\beta,x) = \sup_\alpha[\alpha'(\beta-\overline{b}^m(x)) + \alpha'\alpha/2].$$

Define $\tilde{S}_{ij}^m$, S_{ij}^m in the analogous way, and let the superscript 'o' denote the case where $\overline{b}^m(x)$ is replaced by $\overline{b}(x)$. Theorem 2 and Lemma 3 hold for each m. As $m \to \infty$,

$$\tilde{S}_{ij}^m \to \tilde{S}_{ij}^o, \quad S_{ij}^m \to S_{ij}^o,$$

we have

$$\lim_m \lim_n a_n E_x \tau_J^n = W_J^o - M_J^o(K_i), \tag{5.5}$$

where the limit is uniform for x in a small neighborhood of K_i. Thus, for large enough m, as $t \to \infty$ the path $\{x^n(\cdot)\}$ will spend almost all of its time in a small neighborhood of the set of global minima of $\overline{B}(\cdot)$.

Numerous variations are possible. The $\overline{\xi}_i^m$ can be chosen randomly, but according to some good 'variance reduction' method with the $\overline{\xi}_i^m$ possibly dependent only within a 'cycle'. We could let the cycle length be $m_n \to \infty$, and use $\{\overline{\xi}_i^{m_k}\}$ in the k-th cycle, etc.

References

1. S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE-PAMJ, January 1984.

2. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by simulated annealing," Science 220, May 13 (1983), 621-680.

3. B. Gidas, "Non-stationarity Markov chains and the convergence of the annealing algorithm," J. Statist. Phy., 39, 1985.

4. M. I. Freidlin, "The averaging principle and theorems on large deviations," Russian Math. Surveys, 33, July-December, 1978.

5. S. R. S. Varadhan, Large Deviations and Applications, CBNS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1984.

6. H. J. Kushner, "Robustness and approximation of escape times and large deviations estimates for systems with small noise effects," SIAM J. Appl. Math., February 1984.

7. M. I. Freidlin, A. D. Ventsel, Random Perturbations of Dynamical Systems, Springer 1983.

8. S. Geman, Chii-Ruey Hwang, "Diffusions for global optimization," submitted to SIAM J. on Control and Optimization, preprint, Division of Applied Mathematics, Brown University.

END

FILMED

1-86

DTIC